EFFICIENT DEPICTION OF VIDEO FOR SEMANTIC RETRIEVAL APPLICATIONS BY DIMENSIONALITY REDUCTION OF VISUAL FEATURE SPACE

Amarakoon Mudiyanselage Randitha Ravimal Bandara

(128029E)

Degree of Doctor of Philosophy

Department of Information Technology

University of Moratuwa

Sri Lanka

March 2021

EFFICIENT DEPICTION OF VIDEO FOR SEMANTIC RETRIEVAL APPLICATIONS BY DIMENSIONALITY REDUCTION OF VISUAL FEATURE SPACE

Amarakoon Mudiyanselage Randitha Ravimal Bandara

(128029E)

Thesis submitted in partial fulfilment of the requirements for the degree
Doctor of Philosophy

Department of Information Technology

University of Moratuwa

Sri Lanka

March 2021

DECLARATION

I declare that this is my own work, and this thesis does not incorporate, without acknowledgment, any material previously submitted for a Degree or Diploma in any other University or institute of higher learning, and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Similarly, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic, or other media. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: *UOM Verified Signature* Date: 2/3/2021

The above candidate has carried out research for the PhD thesis under my supervision.

Name of the supervisor I: Dr. Lochandaka Ranathunga

Signature of the supervisor I: *UOM Verified Signature* Date: 2/3/2021

Name of the supervisor II: Associate Professor Dr. Nor Aniza Abdullah

Signature of the supervisor II: *UOM Verified Signature* Date: 2/3/2021

ABSTRACT

The retrieval of temporal digital visual data, either by a text or visual query, requires automatic interpretation, which includes high-level annotation by object detection and recognition for text query-based retrieval and low-level abstraction for visual query-based retrieval. Both the accuracy and the speed of the interpretation become crucial factors in real-world applications, due to the high density of visual data. This study has focused on reducing the complexity of visual data efficiently by dimensionality reduction techniques for the detection and recognition of objects in videos for both textual annotation and visual query-based video frame retrieval. The contribution of the study includes three approaches, i.e., a novel visual feature descriptor based on colour dithering – namely Salient Dither Pattern Feature (SDPF), novel object segmentation method based on the proposed feature descriptor – namely Refining Superpixel and Histogram of oriented optical flow Clustering (RSHC) –, and a novel self-supervised local descriptor – namely Network-in-Network with Restricted Boltzmann Machine (NIN-RBM). The experimental results make it evident that the SDPF is rotation and scale invariant and computationally efficient yet shows similar object recognition accuracy to the state-of-the-art methods with minimum supervision. The results further revealed that RSHC has successfully utilized SDPF for accurately segmenting individual objects by using a very shallow history of motion. Furthermore, according to the results, NIN-RBM has shown the state-of-the-art correspondence matching performance over the existing deep-learned self-supervised binary descriptors, keeping the computation time at the minimum. The overall results support the conclusions that RSHC is capable of accurately segment objects in a video, and then SDPF can be successfully used for recognizing the segmented objects. Moreover, NIN-RBM can be used to reliably and rapidly retrieve video frames related to any visual query. Since NIN-RBM is a local descriptor, it can be further used for locating of high-level objects and estimating their poses precisely, to improve the details of semantics retrieved from video data.

Keywords: dimensionality reduction, colour dithering, deep learning, video segmentation, object recognition, correspondence matching, binary descriptor

ACKNOWLEDGEMENT

First and foremost, I consider it is my bounden duty to record here my sincerest gratitude and appreciation to my supervisor, Dr. Lochandaka Ranathunga for his productive cooperation, guidance, and supervision extended throughout this research project. If not for his encouragement and support, this project would never have been a reality. Similarly, I unreservedly thank my co-supervisor, Associate Professor Dr. N.A. Abdullah from the University of Malaya, Malaysia for guiding me to succeed in this endeavour.

I express my heartfelt thanks to the Vice-Chancellor of the University of Moratuwa, the Dean of the Faculty of Information Technology and the Head of the Department of Information Technology, the non-academic staff members of the Faculty of Information Technology, University of Moratuwa for the opportunity given to commence my research work at the University of Moratuwa and facilitating me to carry out the same successfully. Further, I extend my sincere gratefulness to the Vice-Chancellor of the University of Sri Jayewardenepura, the Dean of the Faculty of Applied Sciences and the Head of the Department of Computer Science, and my colleagues in the academic staff for their assistance in many ways, which was a blessing for completing this research study. Thanks, are also due to the chairman and the staff of the National Research Council, Sri Lanka for granting a research scholarship under grant number NRC/12-017 to financially support this endeavour. Also, I must be thankful to the Senate Research Committee of University of Moratuwa for the financial support provided to carry out this study. Furthermore, I am indebted to the CEO of LK Domain Registry for granting me the Prof. V.K. Samaranayake Research Grant in order to continue my research studies at the University of Malaya, Malaysia.

I declare my salutation and admiration for all the esteemed authors, researchers, and philosophers for their great theories, research, publications, and ideas, which have been an enormous support to enhance this research work.

The support and motivation provided by my postgraduate friends, Mr. K.A.S.H. Kulathilake, Mr. V. Senthooran, Mr. B. Hettige and Mrs. M. Sirisuriya who are affiliated with the Faculty of Information Technology, University of Moratuwa, too deserve mentioning with a debt of gratitude.

I am grateful to my parents, my sister, and brother, for all their encouragement and support extended right throughout this endeavour. If not for them, this project might not have been a reality. Last but not least, it is the dedication of my loving wife and daughter who were bearing all the burdens without passing them to me throughout the past few years that made me complete this research. Finally, I am grateful to all those who assisted me in numerous ways during the course of this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Acronym	Definition
ALOI	Amsterdam Library of Object Images
ANN	Artificial Neural Network
ARM	Advanced RISC (Reduced Instruction Set Computing) Machine
BOW	Bag of Words
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
CCTV	Closed-Circuit Television
CDPC	Compact Dither Pattern Code
CIELAB	International Commission on Illumination. L, A and B are colour components
CIFAR	Canadian Institute for Advanced Research
CKN	Convolutional Kernel Network
CNN	Convolutional Neural Network
CNNH	Convolutional Neural Network with Hashing Layer
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
DBD-MQ	Deep Binary Descriptor with Multi-Quantization
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DDD	Dither Density Descriptor
DNN	Deep Neural Network
DSH	Deep Supervised Hashing

Acronym	Definition
EHD	Edge Histogram Descriptor
FAST	Features from Accelerated Segment Test
FC-GPHOG	Fused Colour-Gabor Pyramidal Histogram of Oriented Gradient
FREAK	Fast Retina Key- point
GAP	Global Average Pooling
GBRBM	Gaussian-Bernoulli Restricted Boltzmann Machine
GPHOG	Gabor Pyramidal Histogram of Oriented Gradient
GPU	Graphic Processing Unit
HOG	Histogram of Oriented Gradients
HOOF	Histogram of Oriented Optical Flow
HSDPF	Hessian based Salient Dither Pattern Feature
HSV	Hue Saturation Value
HTD	Homogeneous Texture Descriptor
IFV	Improved Fisher Vector
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
ITQ	Iterative Quantization
LAP	Local Average Pooling
LDA	Linear Discriminant Analysis
LDB	Local Difference Binary
LPP	Locality Preserving Projection
LSH	Local Sensitivity Hashing
MLP	Multi-Layer Perceptron

Acronym	Definition
NIN	Network in Network
NIN-RBM	Network in Network with Restricted Boltzmann Machine
ORB	Oriented FAST and Rotated BRIEF
PCA	Principle Component Analysis
PHOG	Pyramidal Histogram of Oriented Gradient
RAM	Random Access Memory
RBF	Radial Basis Function
RBM	Restricted Boltzmann Machine
ReLU	Rectified Linear Unit
RFD	Receptive Fields Descriptor
RGB	Red, Green, and Blue
ROC	Receiver Operating Characteristics
RPi	Raspberry Pi
RSHC	Refining Superpixels using HOOF and Colour
SDPF	Salient Dither Pattern Feature
SGD	Stochastic Gradient Descent
SIFT	Scale Invariant Feature Transform
SIMD	Single Instruction Multiple Data
SKAR	Smoothed Keypoint-Matching Ratio
SLIC	Simple Linear Iterative Clustering
SURF	Speeded Up Robust Feature
SVM	Support Vector Machine

Acronym	Definition
TBD	Texture Browsing Descriptor
VGG	Visual Geometry Group